06.2

Working paper

Solvency II Is Not Risk-Based. Could It Be?

Evidence From Non-Life Calibrations.

Sylvestre Frezal

Mars 2017

PARI
PROGRAMME DE RECHERCHE
SUR L'APPRÉHENSION DES RISQUES
ET DES INCERTITUDES

Solvency II Is Not Risk-Based. Could It Be? Evidence From Non-Life Calibrations.

Sylvestre Frezal [1]

Abstract: The aim of Solvency II is to be *risk-based*: capital requirements are considered to be founded on a risk measure. We focus on premium and reserving risks, which represent 40% of capital requirements for non-life insurance companies in Europe, and draw on internal robustness tests to demonstrate that this measure is unreliable. There are three possible explanations for this lack of reliability: a political economy factor, an idiosyncratic factor, and an epistemological barrier. We examine each of these and evaluate their significance, thus providing insights to adapt the design of prudential regulation to these pitfalls.

Keywords: regulation; capital requirements; model evaluation; insurance; Solvency II

JEL Classification: C52, G18, G22

[1] sylvestre.frezal@datastorm.fr, DataStorm, 40 rue Etienne Dolet, 92 240 Malakoff, France. Research affiliate LFA (Laboratoire de Finance Assurance, CREST, Paris, France). Co-director of PARI (Program on the Understanding of Risks and Uncertainties). www.chaire-pari.fr. I would like to express special thanks to Hugues Dastarac for his valuable counsel.

Introduction

From a theoretical point of view, the principle of basing capital requirements on risk measures rather than on volume indicators has been debated, with some authors considering it effective (e.g., Cummins & Phillips, 2009; Eling & Holzmüller, 2008; Weber & Darbellay, 2008) and others taking a more circumspect stance (e.g., Repulo & Martinez-Miera, 2014 regarding banks; Frezal et al., 2016 regarding insurance). This article does not seek to take a direct position in the debate – we begin from the established fact that, in insurance, the Solvency II Directive (EPC, 2009), which redesigned prudential regulation for insurance in Europe, is meant to be *risk-based*:

> The starting point for the adequacy of the quantitative requirements in the insurance sector is the Solvency Capital Requirement (SCR). […] The Solvency Capital Requirement standard formula is intended to reflect the risk profile of most insurance and reinsurance undertakings. (Art. 26)

Where Solvency I was considered too unsophisticated (for example, only distinguishing two sets of lines of business in non-life insurance and weighting the riskiest by a fixed factor of 1.5), Solvency II discerns 12 lines of business, for each of which it calibrates shocks including more than two significant figures (see Appendix 1 for more details).

This refinement of risk characterization is supposed to shed light on two fields. That of regulatory action (licensing and withdrawal), as these measures directly impact the denominator in the solvency ratio. And that of company steering, since these measures directly impact the denominator in the profitability ratio: for insurers and investors seeking to arbitrate between different lines of business, their risk-adjusted profitability should be reflected by these measures. Our goal is to determine whether these refined calibrations are sufficiently robust for the various economic agents to base their operational decisions on them. Indeed, this specific ambition to be risk-based faces three kinds of difficulties:

Scientific. The risk models have yet to prove their robustness. In banking, the FSA observed that banks' internal models diverged by a factor of 6 for a representative portfolio of assets (Samuel & Harrison, 2011). Danielsson (2002, 2008) has also demonstrated that, by retaining a very common range of historical depths and a limited and standard set of underlying mathematical models, we observe up to a two-fold difference between the different possible estimations of a daily VaR at 99% for a liquid vanilla security. In insurance, within the framework of Solvency II, this model risk could be much higher than in the example above. Indeed, (1) the targeted VaR and the data used are not daily, but annual: the lower frequency means that it is even more necessary to extrapolate on the basis of a limited sample size, which degrades the robustness of the statistics that can be estimated, and (2) certain data, such as those derived from the interpretation of flexible internal guidelines, must be systematically preprocessed before being used[2], which increases recourse to expert judgments and thus expands the range of non-illegitimate final results.

Interpretive. Even if the mathematical definition of the risk measure is univocal (an annual VaR at 99.5%), the physical meaning associated with this quantity, as well as the associated calibration methods, are more ambiguous. As such, we do not know whether this probability should be interpreted as "each year, one company in 200 goes bankrupt" or "every two centuries, the entire market goes bankrupt". These two formulations represent polar cases, which, in the first instance would lead to quantifying only specific risk and, in the second, only systemic risk. They constitute frames of mind, as neither the different accounting periods nor the different companies are independent, but privileging one interpretation or another leads to different quantifications – in the one case, concerned with differentiating the behaviors of the different companies, and in the other with preventing the impact of a systemic crisis. Leroy and Planchet (2010) thus interpret a period of upward adjustment of calibrations as the mark of a movement to fight against systemic risk[3].

Political. While calibrations are considered to be technical (simple econometric measures), they also contain political choices. The most flagrant example (which is criticized) is that there is no postulated risk for sovereign debts. For example, Greek public debt was, and continues to be, considered to be risk free: its annual VaR at 99.5% is considered zero.[4]

These three points impose serious limitations to Solvency II's ambition to offer a framework "based on risks", but such qualitative critiques are usually dismissed

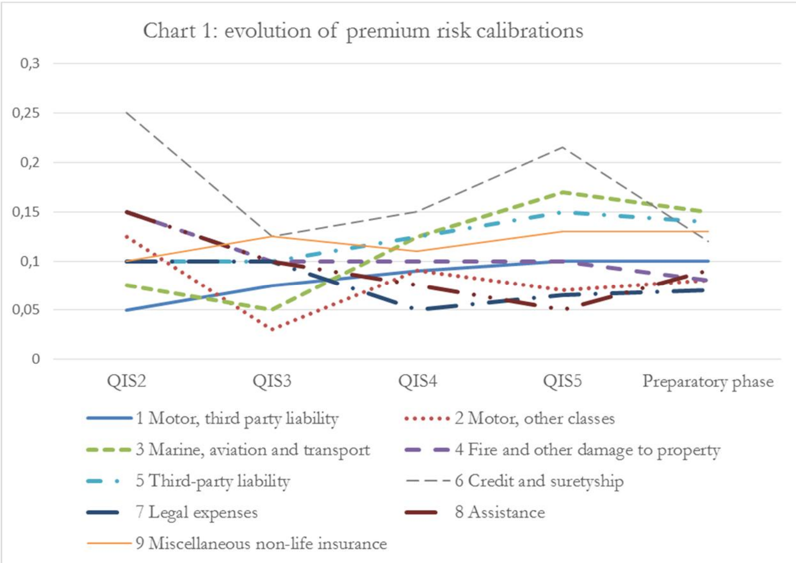
---

[2] For example, the estimate of the ultimate claims necessitates a preprocessing of reserving data provided by the claims service as some data would bias the extrapolations by distorting claims development (due to revisions of reserving policies, atypical IBNRs linked to a wave of claims coming in just before closure, reprocessing of severe claims, etc.).

[3] We can formulate this concrete interpretive ambiguity of VaR and the associated calibration differently by taking for example the probability of the French state going bankrupt. If we consider it impossible that the French state will go bankrupt this year, but not exclude that it might do so in the next 200 years, should we consider the annual VaR at 99.5% to be zero (the organization will not be affected by this risk) or non-zero (the market may be affected by such an event in the next 200 years)? While the mathematical definition appears to be univocal, we find ourselves at an impasse when interpreting the quantity in question, and hence when knowing how to measure said quantity.

[4] It is completely understandable, healthy, and necessary to be concerned about the impacts of a requirement that has massive consequences on the financing of the economy and the stability of a market before endorsing it, but it does not fit with the claim of being risk-based.

with thoughts such as "*don't throw out the baby with the bathwater*", "*it's a proxy, a good proxy*", "*the order of magnitude is right*", or "*it's not perfect, but it's better than nothing*". However, the high volatility observed in the calibrations leads one to wonder whether these responses, which are also qualitative, are well-founded. For example, in 2006, the European regulator (CEIOPS) considered third-party liability to be riskier than transport insurance, then in 2007 decided the contrary (twice less risky), then in 2008 returned to its initial hierarchy (1.5 more risky), reversing the order once more in 2010…then finally considering them equivalent in 2014[5]. More generally, charts 1 and 2 present, for premium and reserving risks (which together represent 40% of the capital requirement for non-life insurance companies[6]), the estimated level of risk for each line of business (LoB) over time, as carried out by the CEIOPS in the context of preparatory work for Solvency II. Each intersection of two lines marks a change in opinion regarding the relative risk of two LoB.

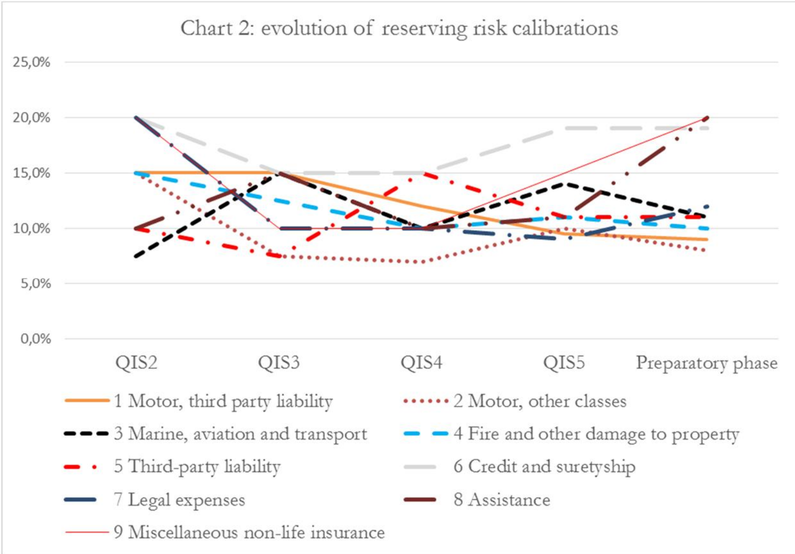


---

[5] This regards reserving risks; we could give many such examples that would also concern premium risk:

- In 2006, assistance was considered 1.5 times riskier than third-party liability (15% vs. 10%); in 2007, it was considered equally risky (10%); in 2010, third-party liability became three times as risky as assistance (15% vs. 5%); and it was finally decided that third-party liability is 1.5 times as risky – about what was estimated in 2008.
- In 2006, credit and suretyship was considered 2.5 times as risky as *miscellaneous* (25% vs. 10%); in 2007, they were deemed equally risky (12.5%); in 2010, credit and suretyship was once more estimated to be riskier (21.5% vs. 13%); and it was finally decided to be less risky (12% vs. 13%).

[6] ACPR, 2011, p.16. Underwriting risks represent 63.2% of the capital requirement for a non-life insurance company, of which 36% correspond to catastrophe risks.



In order to characterize this volatility with a synthetic indicator, we may measure the proportion of LoB couples whose hierarchies remain stable throughout the estimations (cf. Appendix 2). Doing so, we find that in only one-third of the cases (31% for both premium risk and reserving risk) the assessment of one LoB as riskier than another remains stable over the course of the studies. As practitioners review the estimations in such a way, we doubt the ability of this technology to effectively capture risk.

In the first section of this article, we will determine whether these observations should lead us to consider that the calibrations do not provide operationally useful information or whether, on the contrary, they are indeed "better than nothing". In order to do this, we will first set out the methodology employed to evaluate a signal-noise ratio; we will then specify the data used; and finally, we will present the results and their interpretation.

In the second section, we will analyze the different possible causes of this situation, both technical and political, in order to determine the significance of each. This will allow us to propose prudential design improvements.

Section I. Are These Risk Measures "Better than Nothing?"

Danielsson (op. cit.) in the field of banking capital requirements, Planchet & Kamega (2013) and El Karoui et al. (2015) in regard to the economic capital of insurers[7], have tested the impact of changing the set of inputs, thus exploring

[7] Through life reserving, thus by complementarity, the available own funds.

the quantification of error margins. We have developed an alternative methodology, based on the observation and comparison of different official calibration sets devised by regulators and industry actors. It has the advantage of being agnostic as regards "true value" and is thus able to claim objectivity. However, it has the disadvantage of measuring only the internal credibility of a measurement system, not its validity. As such, it cannot confirm the operational pertinence of calibrations, but it does enable us to deny it, when applicable.

Methodology

The aim is to measure the relative amplitudes of noise and signal. In keeping with the expression *signal-noise ratio* used by Sims (2003) and Alesina & Tabellini (2007), signal corresponds to the characteristic amplitude of information that a measure seeks to register and noise to the characteristic magnitude of disturbances that muddle this information.

The relative assessment of noise and signal should enable us to determine whether a measure provides operationally useful information. Let us take an example: suppose we use an altimeter to fly a plane. If, due to the characteristics of the plane at the beginning of the 20th century, it cannot fly higher than 1000m and the altimeter is only accurate to within 10,000m, then the amplitude of the signal we seek is low regarding the interference noise, and the altimeter has no operational utility. Inversely, if the plane can fly up to 10,000m and the altimeter is accurate to within 1000m, it is useful in certain circumstances (not for landing, but at least for avoiding a mountain).

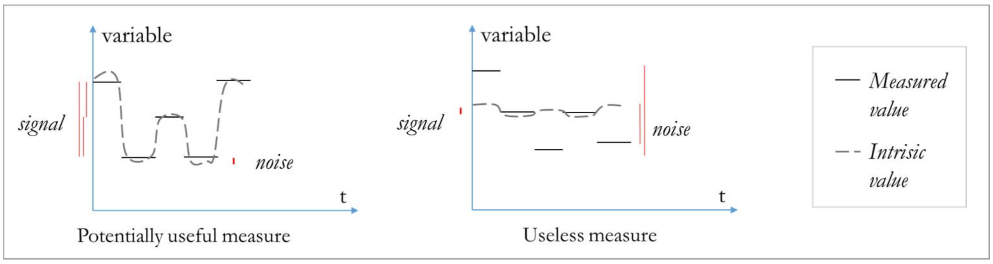


Figure 1: signal-noise visualization of a measure's operational pertinence

To assess the dispersion of measures relating to noise and to signal respectively, we compare (see figure 2):

- a *noise* indicator that reflects the dispersion of alternative calibrations of the same quantity (this is characteristic of the difference between evaluations of *a given risk* assessed through *different approaches*)

- to a *raw signal* indicator that captures the dispersion of calibrations for different quantities (this is characteristic of the difference between evaluations of *different risks* by *a given approach*).

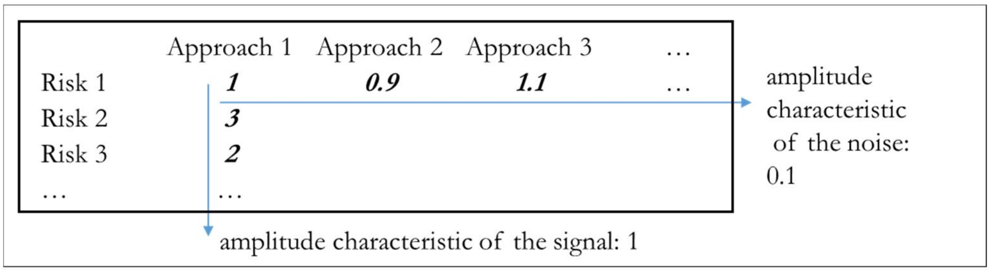


Figure 2: Illustration of signal and noise

These dispersions are measured for a set of risks and a set of approaches with the aid of two common and intuitive indicators:

- the standard deviation
- and for the "extreme" dispersion, the ratio $\frac{max-min}{min}$.

This second indicator can only be positive, and if it is greater than 1, this indicates there is more than a two-fold difference between the lowest and highest estimates.

From these indicators, we create the adjusted signal-noise ratio (SNR_a) as follows:

$$SNR_a = \frac{raw\ signal\ - noise}{noise}$$

Using such a ratio rather than a simple signal-noise ratio enables us to take into account the fact that the raw signal indicator, corresponding to the measure of dispersion within a given approach, is itself affected by noise. Thus, our adjusted indicator is such that if all the cells in the table (figure 2) are uniformly filled by a random generator, the raw signal and noise indicators being asymptotically identical, then the expected value of the signal-noise ratio would be zero, which reflects the fact that the perceived raw signal would just be noise without information[8]. This adjustment is based on the hypothesis (formalized and discussed in Appendix 3) that the error matrix is isotropic[9].

Data

What different kinds of approach can be used to measure a given risk? We will look separately at two categories of approach: one being those approaches developed by the European regulator to calibrate the *standard formula*, the other being those developed by a large insurer for its internal model.

QIS

Solvency II was subjected to a series of quantitative impact studies (QIS) designed to measure the capacity of organizations to perform calculations, to assess the impact of this regulation's implementation on their need for financing, and to refine calibrations. Four of these studies[10] proposed a calibration of the 1-year 99.5% VaR calibrations for a series of risks, and the final implementing measures led to a final set of calibrations. As such, we have five sets of risk measures at our disposal, each representing a "possible approach" determined by the CEIOPS or the European Commission.

The conformism (generated, e.g., by market standards[11]) biases the analysis by restricting the range of calibrations in relation to the set of those that are "scientifically legitimate". The observed dispersion between approaches is thus underestimated, generating an artificial convergence of results and thus favoring an image of the system as possessing internal credibility. To limit this bias, we do not consider calibrations of market risks, where the use of VaR has long been widespread and thus methods have had more occasion to be standardized. Among underwriting risks, we focus on premium and reserving risks – those for which we have a significant number of standardized stresses, which allows us to

[8] whereas if we retained a raw ratio (raw signal to noise) in this case, we would obtain a misleading result of 1, implying that there is a signal of the same amplitude as the noise.
[9] Given the figures, should this hypothesis not be true, it would not alter the nature of the results obtained.
[10] QIS 2, 3, 4, and 5. QIS 1 focused on the valuation of the balance sheet and included neither capital requirement tests nor calibrations of related requirements.
[11] The directive explicitly emphasizes the necessity that, in terms of calibration, the "specification is in line with generally accepted market practice" (EPC, 2009, Art. 122.4). To give an example in terms of implementation, the *QIS 3 calibration paper* (CEIOPS, 2007b, p. 12) explains that for the purpose of estimating VaR at 99.5% for non-life premium risk, a value of 3σ is retained, "*assuming a lognormal distribution of the underlying risk*". This hypothesis applies to every LoB considered and thus tends to guide and confine the choices of laws selected from all of the possible extrapolations.

conduct our comparative study. We thus measure the risks of the 9 LoBs of direct insurance and proportional reinsurance.

An internal model. Those insurance companies that prefer to use an internal model may opt to do so. We have at our disposal calibrations measured internally by an insurance group, at different dates and for each of the risks considered. The group in question is the French sub-group of one of the four global European insurance groups, which developed an internal model and is present in each of the LoB.

We have usable data for reserving risk[12]:

- on the five main LoB (motor, third-party liability; motor, other classes; marine, aviation and transport; damages; and third-party liability),
- over six years, from the end of 2009 to the end of 2014. The insurer had developed a model over several years preceding 2009. The methodologies thus had time to crystallize before the data range we have here, which is likely to create an artificial stability. We therefore certainly underestimate the margin of error more significantly than with QIS data.

Results

QIS

Figures 15 and 16 (Appendix 4) present the set of calibrations[13]. Figure 3 synthesizes these results, presenting the averages of the noise indicators (average across LoBs of the dispersion between approaches) and signal indicators (average across QIS of the dispersion between LoBs).

[12] The methodology for premium risk differs from that of the standard formula in that it focuses directly on the extreme quantile and not on the difference between the extreme quantile and the best estimate. The interpretation of the risk measure thus does not correspond to a dispersion, and comparisons between classes are no longer directly valid, as each one has a different expected value. We do not have access to these expected values to provide us with comparable "volatility" indicators.

[13] The last two rows show, for each QIS, the value of indicators of dispersion between classes – they synthesize an order of magnitude for the measure's dispersion capacity, which is to say the amplitude of the information supplied. The last two columns show, for each class, the value of indicators of dispersion between QIS – they synthesize an estimation of the amplitude of the noise interfering with the measure. The second indicator, for example, shows that for two-thirds of the LoBs, the calibration changes by a factor of two or more depending on the measure.

	st. dev.		max/min	
	Premium	Reserving	Premium	Reserving
Raw signal	4.10%	3.70%	2.7	1.3
Noise	3.20%	3.40%	1.4	0.9
SNRa	28%	12%	99%	47%

Figure 3: average indicators of raw signal, noise, and adjusted signal-noise ratio (QIS)

The following elements appear to stand out:

- On the basis of the first dispersion indicator (standard deviation), the raw signal essentially captures noise. For the premium risk measure, the adjusted signal to noise amplitude is less than 30%. For the reserving risk measure, this signal-noise ratio becomes utterly negligible.
- The results obtained from the second dispersion indicator (max-min/min ratio), though not as pronounced, tend to confirm this first observation. Thus, for premium risk, the signal-noise ratio is 1 and for reserving risk it is 0.5.

It appears difficult, then, to consider that Solvency II's proposed risk measures offer a significantly higher dispersion than that of an outside interference (be it technical or political).

Internal model

For non-technical reasons, we could expect the calibrations resulting from an internal model to be significantly more stable than those resulting from the standard formula's different tests: first, because of the desired stability of a methodological framework that is to be submitted to a supervisor's approval; second, due to their relative immunization to the fluctuations of political orientations to which European text drafts are subjected; and third, because the data we have been able to use originate several years after the establishment of an internal model and the production of associated figures by the entity in question, allowing for an initial crystallization. As such, the perceived margin of error could be very low for these data.

Nevertheless, reproducing the same studies leads to results that, even if they are indeed slightly better, still show internal noise to remain considerable:

	st. dev.	max-min/min
Raw signal	10.9%	1.8
Noise	6.5%	0.9
SRNa	68%	106%

Figure 4: average dispersion of signal and noise using an internal model (reserving risk, 2009-2014)

We observe that:

- on average, the estimation of risk associated with a given class changes by a factor of 2 over the period (max-min/min=0.9),
- the amplitude of the internal signal is less than or comparable to that of the internal noise.

This appears to suggest that these measures do not provide reliable operational information.[14]

Interpretation: is the "noise" truly noise?

As we mentioned in the introduction, several sources of diversity are likely to make calibrations vary from one exercise to another:

- a change in political trade-off, related for example to pressure exerted by an industry, or to a member state's desire to not put "its" actors at a disadvantage, or to the fear of an overly heightened calibration's consequences on the market, etc. These political assessments were encouraged, potentially, by the fact that the deterioration of the financial market at the time added a strain to companies' capital positions and their ability to raise funds should the new regulation require it.
- a technical change[15], either (i) related to the data themselves, for example on the occasion of a refinement of the data, an increase in historical depth or, on the contrary, a change in products that renders a past calibration obsolete, or (ii) related to expert judgments, for example due to the use of a new kind of model (e.g., a change in the underlying distribution function), or simply a change of personnel responsible for the calculation[16].

[14] The same holds from an ordinal point of view, as the stability indicator for the hierarchy of the five classes over the 6-year period considered is 40%. This means that a judgment on the relative risk serving as the basis for a decision will have been inversed in most cases.

[15] Or a change in the interpretation of the risk measure.

[16] As an example concerning this last point, a manager who knows his team can generally tell which team member has conducted a statistical analysis by observing which "outlying" points have been excluded from the analysis.

The reason for this volatility, be it technical or political, is of little importance. If it is political, this obviously distorts the analysis' purely technical credibility. If the reason is technical, it seems unjustifiable to consider that the insurers' data and the field of actuarial science made a qualitative jump sometime between 2006 (the date of QIS 2) and 2010 (the date of QIS 5) and that, as opposed to 2006, 2010 or 2014 would represent the year when the actuarial science and the available data reached their full realization.

Two arguments may yet explain a dispersion that might be wrongly interpreted as noise:

- The first argument is that risks change and thus it is legitimate that calibrations adapt to them. Our analysis does not permit us to rule out this argument. However, (1) the period of revision for the calibrations observed (which is annual for the first few QIS) is far smaller than the duration of the liabilities under consideration and the associated risks (approximately 5 years on average), and (2) if prior changes to the calibrations did in fact reflect the changes in the underlying risk, their crystallization would generate a gap between the calibration and the underlying risk of a comparable order of magnitude in the future. And there are no plans to revise the directive's calibrations on a regular basis.
- The second argument is that certain of the calibrations from the first few QIS were based on data from some national markets within the European Union (CEIOPS, 2010b, p.189), and thus the expansion of the database could explain the variations. Once again, the quantitative pertinence of this argument cannot be discounted by our analysis. However, if this were to be the source of the variations, it would mean that the heterogeneity of the underlying risks among national markets is such that imposing uniform measures generates a gap between the underlying risk and its measure at an order of magnitude comparable to that of the "noise" measured. Thus, it is effectively noise that we find when we attempt to determine the risk profile of a given company, which is the objective and claim of Solvency II.

As such, in the context of the Solvency II Directive, whose calibrations are set to be fixed and homogeneous, the arguments above do not call into question the consideration of the dispersion between the different QIS (and with the Level 2 implementing measures) as a matter of noise – of an undue discrepancy between a company's capital requirement and its risk[17]. This leads to a validation of the interpretation of instability as an unreliability. Is it possible to identify the sources of this noise and to adapt the design of prudential regulation?

Section II: The Causes

In this second section, we study three explanations that may account for the observed instability of the calibrations: the possible disruption of technical calibrations by political issues, the importance of the idiosyncratic factor, and a potential epistemological barrier. To this end, we will systematically rely on the data and the methodological elements provided by the CEIOPS to support its proposed QIS calibrations.

Political economy?

It is possible to reconstitute the aggregate level of calibrations associated with the premium and reserving risks for a company representative of the market by using the calibrations for each LoB provided by the CEIOPS. Here, we consider a company whose relative weights in each LoB correspond to those of the French market. Figure 7 shows this development over time.

	2006 QIS2	2007 QIS3	2008 QIS4	2010 QIS5	2014 Prep. phase
premium	12.7%	7.7%	9.9%	9.9%	9.2%
reserving	13.2%	10.5%	11.8%	10.7%	10.3%

Figure 7: Estimation of the overall net risk calibration, representative company

The comparison of these developments with the macroeconomic context opens up the hypothesis of a political steering of the average level of stress according to the following chronology:

[17] As concerns internal model, the sources of variations both over time and in regard to the standard model may be of a political or a technical nature, or both. If they are political in nature, then it is a different political nature than that of the standard formula: here, this would be the company's internal issues (e.g., supporting a strategic choice or steering an internal message towards the holding company or external actors such as rating agencies). If such dispersion is related to expert choices, this destroys the very purpose of a scientifically established risk-based perspective; if it is related to a change in the data, this contradicts the notion that a standard formula (stable over time) could be pertinent. Moreover, in this case, it appears that the period of obsolescence for the information provided by these calibrations is lower than the duration of the portfolio, which raises doubts as to the pertinence of their use in steering by companies.
Once more, the reasons for the dispersion of parameters deemed to be technically pertinent is of no matter; the dispersion of observations is to be classified as noise and lead to a disqualification of these calibrations as reliable tools for steering and regulation.

- an initial calibration (2006) was carried out "blind" in regard to its impacts,
- a downward revision of this first calibration, judged to be too high and thus too detrimental to the market (2006-2007 development),
- followed by an upward correction in a context of financial crisis leading to the strengthening of prudential vigilance and the associated quantitative constraints (2007-2008 development),
- then a halting stabilization during negotiations between the CEIOPS, industry, and the States (2008-2010 period),
- followed by one last political reduction on the occasion of the Commission's final choice of parameters in 2014.

If not this exact chronology, can we at least validate the hypothesis that political choices are significant sources of variation in calibrations?

Data used

The CEIOPS has occasionally accompanied QIS calibrations with supporting evidence:

- No evidence was provided for QIS 2 (CEIOPS, 2006).
- Almost no evidence was provided for QIS 3 (CEIOPS, 2007b). However, some information suggests that this new set of calibrations fits in a global context of expectations of stress reduction.[18]
- Supporting evidence is nearly inexistent for QIS 4, with the CEIOPS indicating, for example, that "*attention was paid to the industry's QIS 3 feedback regarding the calibration of the SCR formula, but being globally pleased with the QIS 3 calibration, CEIOPS decided not to substantially challenge the QIS 3 calibration*" (CEIOPS, 2008).
- However, in 2009 (an abortive calibration considered at the time as the CEIOPS' final proposition to the European Commission, which did not give rise to an impact study) and in 2010 (QIS 5, effectively the CEIOPS' final recommendations to the European Commission), the CEIOPS provided information about the calibrations and the methodology employed (CEIOPS, 2009 & 2010b).

[18] For example, between QIS 2 and QIS 3, the closed formula enabling the transformation of the volatility measure into a regulatory risk measure shifted from a formula reflecting a TVaR at 99% to one reflecting a VaR at 99.5% (both based on the same premise of an underlying log-normal distribution), thereby generating a decrease of 5 to 10% in the regulatory calibration for an unchanged volatility.

For each of these two exercises,

(i) the CEIOPS applied four estimation methods for premium risk and six for reserving risk to the data at its disposal (a few countries in 2009, and a larger perimeter in 2010). All of the methods used were based on the same assumption of underlying law and differed by the number of degrees of freedom retained.[19]

(ii) Then, risk by risk, the CEIOPS identified which method(s) seemed the best fit(s),

(iii) and chose (a) certain method(s) to determine its proposed calibration of the standard deviation for each of the risks (sometimes the average of the methods retained, sometimes a figure deemed "consistent" with the adopted methods).

Results

The set of methods adopted for the calibration proposal (iii) regularly deviates from the set of methods considered most pertinent from a technical standpoint (ii). Focusing on premium risk, LoB by LoB, figure 8 synthesizes the set of methods considered most pertinent and the set of methods adopted for the final calibration: there is no LoB for which these two sets are identical.

	CP 2009		QIS 5	
Premium risk	best fit	retained	best fit	retained
1 Motor, third party liability	3-4	1 - QIS4	4	1-4
2 Motor, other classes	3-4	3 - 1 -QIS 4	2-4	1-2-4
3 Marine, aviation and transport	2-3	4-2	2-4	1-2-4
4 Fire and other damage to property	2	1 - QIS4	none	1-2-3-4
5 Third-party liability	3	?	4	1-4
6 Credit and suretyship	2	?	2-4	1-2
7 Legal expenses	2-4	?	2-4	1-2
8 Assistance	2-4	2 - QIS 4	2	1-2
9 Miscellaneous non-life insurance	2-4	2 - QIS 4	2-4	1-2-4

Figure 8: methods considered optimal and methods adopted for the calibration

The same phenomenon can be observed in regard to reserving risk. For example, the consultation paper (CP, 2009) identifies three "pertinent" estimations of

[19] e.g., (1) the calculation of a unique expected value and a unique volatility, (2) the calculation of a unique expected value and of a volatility per company (to be averaged), (3) the calculation of an expected value per company (to be averaged) and of a unique volatility, etc.

25%, 23%, and 45% for LoB 9, then concludes without justification that the calibration should be 20%.

Generally speaking, the justification for this deviation (passage from step (ii) to step (iii)) is often very weak, if not non-existent. For example:

> "Overall conclusions: Method 4 and 2 provide a good fit. This would imply a factor of 14% on average based on the fitted results. Method 5 does not allow for diversification and in views of the graph above seems to ignore some important observations. A final factor considering method 5 and method 1 has been selected." (CP2009, reserving risk, LoB 7).

It is possible to estimate the total impact of this divergence between the "technical optimum" and the adopted proposal. Figure 9 compares, for our representative company, the average estimation of risk according to the methods judged to be pertinent to that which results from the final calibration proposal. The proposed calibrations result in a significantly weaker estimate than the one to which the purely technical calibrations would have led.

Weighted average	Best fit	Proposal
QIS 5	15.1%	13.0%
CP 2009	22.8%	*16.3%*
Reminder of QIS 4		13.9%

Figure 9: LoB volume-weighted averages of standard deviations linked to premium risk (gross of reinsurance), estimated by the CEIOPS

It is clear here that the successive calibrations took into account a significant non-technical dimension[20]. Does this mean that if the calibrations are not reliable as measures of risk, it is due solely to the interference of political arbitrations distorting the math? Would improving the independence of technical authorities be enough to restore technical credibility to this regulation that claims to be risk-based?

An idiosyncrasy?

An explanation for the dispersion of calibrations observed between different QIS may be the level of indiosycrasy on the market. Indeed, let us recall that the first couple QIS calibrations were carried out with limited data, sometimes

[20] A paradoxical picture emerges from the above-mentioned elements: a technical authority, the CEIOPS, which initially allowed itself to forgo justifying its stance and thus appeared self-sufficient, later, upon being asked for a technical opinion intended to inform a decision, issued recommendations with a patently political component.

restrained to one national market. When firms are very heterogeneous, it is normal for a modification of the calibration sample to alter the parameter.

Certain calibration methods led the CEIOPS to estimate a VaR for each of the companies whose data it had obtained. Figures 5 and 6 show, for those of these methods deemed pertinent by the CEIOPS, the first and third quartiles of premium risk (figure 5) and reserving risk (figure 6) estimated for each LoB.

Premium	Method	75th percentile	25th percentile	max/min
2 Motor, other classes	2	18%	8%	2.3
3 Marine, aviation and transport	2	109%	28%	3.9
4 Fire and other damage to property	2	61%	16%	3.8
4 Fire and other damage to property	3	96%	25%	3.8
6 Credit and suretyship	2	124%	40%	3.1
7 Legal expenses	2	27%	11%	2.5
8 Assistance	2	14%	6%	2.3
9 Miscellaneous non-life insurance	2	77%	15%	5.1

Figure 5: dispersion of estimated volatility between the first and third quartile of firms, according to the methods adopted for the final choice and determining one volatility per company (premium risk, QIS 5)

Reserve	Method	75th percentile	25th percentile	max/min
1 Motor, third party liability	1	17%	6%	2.8
1 Motor, third party liability	2	40%	10%	4.0
2 Motor, other classes	1	40%	14%	2.9
3 Marine, aviation and transport	1	63%	32%	2.0
3 Marine, aviation and transport	2	365%	43%	8.5
4 Fire and other damage to property	1	40%	13%	3.1
4 Fire and other damage to property	2	81%	24%	3.4
5 Third-party liability	1	50%	13%	3.8
6 Credit and suretyship	1	81%	29%	2.8
8 Assistance	1	87%	29%	3.0
9 Miscellaneous non-life insurance	1	72%	25%	2.9

Figure 6: dispersion of estimated volatility between the first and third quartile of firms, according to the methods adopted for the final choice and determining one volatility per company (reserving risk, QIS 5)

We see that the estimated risk level differs by a factor of 3.5 on average between the first and the third quartile (it varies by a factor of 2 to 8 depending on LoB and method).

This dispersion of risk measures between companies within the same LoB is much higher than the dispersion between LoBs that Solvency II claims to capture (for premium risk, the class with the highest risk (MAT, 15%) differs from that with the lowest risk (legal expenses, 7%) by a factor of 2; and by a factor of 2.5 for reserving risk). As such, the choice of tacking a standard formula – calibrated by LoB but homogeneous between firms – onto companies (that are so heterogeneous within the same LoB) provides no significant added value in terms of risk measurement compared to a system that treats classes homogeneously[21].

Therefore, to be risk based seems to require to generalize the practice of undertaking specific parameters (USP) that aim to capture the risk specific to each company. However, if this were necessary to achieve Solvency II's objective, would that be enough?

An epistemological barrier?

Independent of the political interference and the idiosyncrasy to which we have just called attention, the analyses provided by the CEIOPS to support their calibration proposals show that, using a constant dataset, different methods lead to widely dispersed results[22]. Figure 10 highlights the dispersion of the calibrations (ratio between the highest and the lowest) for each LoB according to the method utilized: depending on the LoB and the exercise considered, the gap between estimates varies by a factor of 2 to 60.

[21] As such, the instability of the measures is partly the characteristically transitory and halting consequence of the progressive implementation of the new system, without this calling into question the "noisy" character of these developments.

[22] At the aggregate level, figure 12 shows for our representative diversified company that if a single method had been chosen to assess the risk corresponding to all lines of business transversally, we would have obtained a premium risk estimate up to 4 times higher (and a reserving risk estimate up to 10 times higher) depending on the method.

weighted averages	Method 1	Method 2	Method 3	Method 4	Method 5	Method 6
Premium - CP 2009	*11%*	*9%*	*25%*	*39%*		
Premium - QIS 5	10%	12%	24%	19%		
Reserving - CP 2009	19%	22%	48%	6%	17%	31%
Reserving - QIS 5	16%	19%	46%	4%	15%	31%

Figure 12: estimate of the standard deviation (gross of reinsurance) connected to premium and reserving risks for a representative company according to the different methods employed

ratio max / min	Premium		Reserves	
	CP2009	QIS 5	CP2009	QIS 5
1 Motor, third party liability	2.4	4.2	7	13
2 Motor, other classes	2.2	2.7	5	7
3 Marine, aviation and transport	3.3	3.1	7	7
4 Fire and other damage to property	2.8	1.8	6	11
5 Third-party liability	2.6	1.8	23	22
6 Credit and suretyship	6.0	3.2	5	60
7 Legal expenses	8.0	17.3	12	16
8 Assistance	6.0	5.5	34	26
9 Miscellaneous non-life insurance	48.0	4.1	2	20

Figure 10: ratio between extreme estimates provided by the different methods used

Even when restricted to the only methods adopted by the CEIOPS on a case-by-case basis to determine the final calibrations, the dispersion remains very high – up to a factor of 8 (cf. figure 11) – much higher than the dispersion of calibrations between classes that was retained in the end[23].

ratio max / min	Premium		Reserves	
	CP2009	QIS 5	CP2009	QIS 5
1 Motor, third party liability	1.3	2.8	1.5	4.2
2 Motor, other classes	1.7	1.6	2.5	2.2
3 Marine, aviation and transport	1.7	1.4		3.0
4 Fire and other damage to property	1.3	1.8	1.8	2.6
5 Third-party liability		1.6	1.1	3.1
6 Credit and suretyship		1.2		1.0
7 Legal expenses		1.7	4.0	3.5
8 Assistance	1.3	1.3	7.3	3.9
9 Miscellaneous non-life insurance	7.8	2.2		1.5

Figure 11: dispersion between the results obtained by different methods and taken into account for the final choice

Given the current state of knowledge[24], Solvency II's aim to build a risk-based system appears utopian, since the margin of error associated with the risk

23 Let us also highlight that this measure underestimates the "model's margin of error". Indeed, the only source of variation between the different methods adopted is, as indicated earlier, the number of degrees of freedom; but plenty of other parameters (such as the historical depth deemed pertinent or the choice of the underlying law of distribution enabling to transform volatility into regulatory VaR) are fixed here though they too could be subjects of diverging statistical assessments.

24 One issue, related to the fact that the methods used by the CEIOPS are agnostic in regard to the idiosyncrasy discussed above, will have to be considered in the longer term: is the barrier that we have just highlighted itself a result of this agnosticity and therefore of the absence of a

estimations is much greater than the dispersion of risks between them. As such, it would seem more reasonable to build an *uncertainty-based* prudential system founded on a qualitative, fixed assessment of the uncertainty of each LoB rather than a risk-based one founded on a vain ambition of quantitative measurement.

Conclusion

One can discuss whether or not, in a context where issues concerning the protection of policyholders coexist with those concerning the development of insurance markets and the financing of the economy, the financial requirements imposed on firms should be based on prudential criteria alone. One could then discuss whether or not, from a strictly prudential standpoint, it is fully appropriate to base capital requirements on a risk measure. In any case, this was the double choice made by Solvency II. But it appears that the objective was not met: the calibrations on which capital requirements are based are not reliable. Let us stress that this study suggests Solvency II paradoxically led to a deterioration of resource allocation. Indeed, the allocation of capital resources has not been improved, and, in view of the energy expended on its conception and deployment, the allocation of human capital resources to which it has led is suboptimal.

Beyond this, one could question whether such a risk-based ambition can even be realized. We can't consider that risk measurement technologies don't currently exist, but the evidence presented in this article demonstrates that the issue is not a question of *refining* the precision – rather, it is a question of the current *inability* to characterize the relative risk of one activity in comparison to another. As such, we cannot rely on this basis for decision-making.

In this context, what should be done regarding the technical limitations and political trade-off issues?

Better identification of technical limitations

The study we have conducted can be enriched in two ways. First, by broadening the scope of the analysis from only premium and reserving risks to all risks. Second, by not only testing internal credibility but also the calibrations' intrinsic margin of error through several sets of alternative inputs. While some studies of

methodological framework concerning the resulting idiosyncrasy? In this case, the barrier could be crossed. If not, does it constitute a technically insurmountable hurdle? To answer this question would doubtless require a systematic exploration of the sources of the margins of error and their impact.

sensitivity to hypotheses have been conducted[25], it would now be useful to review these analyses in order to determine the associated margin of error, for each capital requirement component. We insist that the issue at stake is not so much sensitivities – which are forgotten as soon as they are presented and often serve as a discharge to present and use an unreliable result – as it is the *margin of error*; the sensitivity to a parameter should only be a means of determining it.

Conscientious steering

Precisely, the sensitivity studies that were conducted were, to our knowledge, never meant to determine whether the measures were sufficiently precise to serve as a basis for decision-making. Still, they were intended to enable one hypothesis to be chosen over another. Three reasons can be put forward to explain this: first, the fact that a risk-based system was going to be implemented, so the issue was its characteristics rather than its merits; second, the modelers may have wanted to avoid devaluing their own work and thus risk being seen as useless[26]; third, the diffuse and omnipresent feeling that it was "better than nothing" – that the result of a calculation is superior *in principle* to a fixed ratio.

We have shown that such a sentiment is unfounded at this stage. In the current state of knowledge, an insurance company's capital requirement under Solvency II is the aggregation of several dozen components, for which it is not possible to know whether the relative orders of magnitude are being respected. The result is thus completely devoid of meaning: the implications of this lesson for current steering should be noted on both the macro-economic and the micro-economic levels.

At the macro-economic level, when these calibrations appear to be technically unreliable, they are no longer legitimate for establishing insurers' capital requirements. This calls for an effort to rethink these requirements at a political level to articulate the whole set of impacted dimensions (e.g., financing of the economy, development of a market, etc.) beyond an artificial risk-based

---

[25] And while they may be scattered, they are numerous (both internally and at aggregated levels for professional federations and regulators): the impact of the choice of such or such a yield curve reference, of the choice of such or such a calibration of market stress, of the choice of such or such a cat model rather than another, of such a management rule instead of another, of such a reference to ratings or to market spreads so as to evaluate a portfolio's sovereign default risk, etc.

[26] Questioned by a consultant for a survey on ALM practices and strategic asset allocation methods, I had the following exchange:

- Me: The important thing is to know and to relate the margins of error.
- Him: Yes, you have to present the sensitivity analyses.
- Me: No, the sensitivities are anecdotal. What's important is the margin of error.
- Him: But if we present the margins of error, the models will lose their credibility!

calibration. Of course, we could consider that the political issues have been taken into account, as the CEIOPS seems to have censured itself by downward biasing its recommendations with respect to its technical estimations. But the CEIOPS, whose expertise is exclusively technical[27], does not have the legitimacy to make decisions and, indeed, does not seem to take responsibility for it. Nothing, to our knowledge, suggests that these choices have been steered conscientiously on the basis of a structured analysis.

At the micro-economic level, the main danger is for these tools to effectively be used as risk indicators in steering companies. The most precarious firms, thus those for which the quality of risk management is most sensitive, are all the more exposed because the pressure of a low solvency ratio will lead them to align their management with this framework in order to optimize the image they give to the regulator. Both in terms of discourse and implementation, with regard to use test requirements and risk management issues, it is important to dissociate the apprehension of risks from Solvency II's quantifications as much as possible.

Bibliography

ACPR (Autorité de Contrôle Prudentiel et de Résolution), "Solvabilité 2 : principaux enseignements de la cinquième étude quantitative d'impact (QIS5) ", *Analyses et synthèses* n°1, 24p, 2011

Alesina, A. and Tabellini, G., "Bureaucrats or Politicians? Part I: A Single Policy Task". *American Economic Review*, 97(1), pp.169-179, 2007

CEIOPS (Committee of European Insurance and Occupational Pensions Supervisors), "Quantitative Impact Study 2 Technical Specification*", CEIOPS-PI-08/06,* 66p., 2006

CEIOPS, "QIS3 Technical specifications, Part I: Instructions", *CEIOPS-FS-11/07*, 119p., 2007

[27] We will not get into the institutional sociology of Solvency II here. Of course, formally, the CEIOPS's calibrations are only recommendations to the European Commission, but they are followed in practice due to the former's technical legitimacy relative to the latter. The weight of the CEIOPS is reinforced by the scope of its recommendations, in the face of which the national Treasuries are in a position to either "comply or explain", thereby inversing the positions of the technical judging authority responsible for applying the rules and the political authority responsible for setting the framework.

CEIOPS, "QIS3 Calibration of the underwriting risk, market risk and MCR", *CEIOPS- FS-14/07*, 53p., 2007b

CEIOPS, "Draft CEIOPS' advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Calibration of Non-Life Underwriting Risk", *CEIOPS-CP-71-09*, 118p., 2009

CEIOPS, "Solvency II calibration paper", *CEIOPS-SEC-40-10*, 384p., 2010

Cummins, J., & Phillips, R., "Capital adequacy and insurance risk-based capital systems", *Journal of Insurance Regulation*, 28(1), pp. 25-72, 2009, Fall 2009

Danielsson J., "The emperor has no clothes: Limits to risk modelling", *Journal of Banking and Finance*, 26(7), pp. 1273-1296, 07.2002

Danielsson J., "Blame the models", *Journal of Financial stability,* 4 (4). pp. 321-328, 06.2008

El Karoui, N., Loisel, S., Prigent, J.-L., & Vedani, J., "Market inconsistencies of the market-consistent European life insurance economic valuations: pitfalls and practical solutions", *working paper*, online publication 2015, https://hal.archives-ouvertes.fr/hal-01242023

Eling, M., & Holzmüller, I., "An overview and comparison of risk-based capital standards", *Journal of Insurance Regulation*, 26(4), 31-60, 06.2008

EC (European commission), "Technical Specifications QIS4", *MARKT/2505/08*, 293p., 2008

EC (European Commission), "QIS5 Technical Specifications", 330p., 2010

EPC (European Parliament and Council), "Directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II)", *OJ L 335*, p. 1–155, 17.12.2009

Frezal, S., “de quoi Solvabilité 2 est-il le nom ?” *working paper PARI*, http://www.chaire-pari.fr, 56p., 03. 2016

Frezal, S., Haguet, E. et Nou, V., “Procyclicité et contrats avec PB : inéluctabilité, avantage et inconvénient” *working paper PARI*, http://www.chaire-pari.fr, 31p., 05. 2016

Harrison, M. & Samuel, S., “Two hundreds millions inputs. Can you trust risk weightings at European banks? ”, *Barclays Capital Equity Research,* 06.04.2011

Kamega, A. & Planchet F., “Construction de tables de mortalité prospectives sur un groupe restreint : mesure du risque d'estimation, Bulletin français d’Actuariat”, *Bulletin Français d’Actuariat*, 13(25), pp.5-33, 01-06.2013

Leroy, G. & Planchet, F., “Que signifie la ruine dans Solvabilité 2 ?”, *La Tribune de l’assurance*, 147, pp. 55-56, 05.2010

Repulo & Martinez-Miera,, “Comparing Flat and Risk-based Capital Requirements”, in *3rd European Banking Authority Policy Research Workshop*, London, 11. 2014

Sims, C., “Implications of rational inattention”, *Journal of Monetary Economics*, 50(3), pp.665–690, 04.2003

Weber, R., & Darbellay, A., “The regulatory use of credit ratings in bank capital requirement regulations”, *Journal of Banking Regulation*, 10(1), pp. 1-16, 11.2008

Appendix 1: Comparative overview of the quantitative requirements of Solvency I and Solvency II

The diagram below summarizes how, in comparison to Solvency I, Solvency II introduces a risk-based calculation of associated capital requirements for the most important balance sheet items[28].

Solvency I	Solvency II

[28] See Frezal (2016) for a more general presentation of Solvency II.

Non-life	
▪ Based on the maximum of one rule linked to stock and one linked to flow ▪ 2 activity categories (max. deviation: 1.5) ▪ Diversification only taken into account by size effect (1 threshold)	▪ Based on the aggregation of several rules linked to stock and flow ▪ 12 activity categories (max. deviation: factor of 2 to 2.5[29]) ▪ Size effect not taken into account, but diversification between classes is (correlation matrices)
Life	
▪ 2 classes depending on whether capital is guaranteed or not (deviation factor of 4) ▪ Fixed add-on for death benefits	▪ Based on point-by-point modeling (taken into account by guaranteed rates, profit-sharing rates, policyholder characteristics) and ALM hedging
Assets	
▪ No impact on capital requirements ▪ Rules limit concentration	▪ Capital requirements depend on class (nature), rating, duration, etc.
Correctives	
▪ Life: available margin decreased if ALM hedging is ill-adapted to guaranteed rates – but no impact on capital requirement ▪ Non-life: reinsurance taken into account *in proportion* to its past impact	▪ Treaty-by-treaty simulation of the impact of reinsurance on extreme events ▪ Possibility to develop internal calibrations (USP: Undertaking specific parameters) or internal models ▪ Diversification, etc.

Figure 13: Solvency II, or the introduction of a risk-based capital requirement

Appendix 2: Stability measure of the hierarchy of risks

Aiming to improve the management of risks and decision-making processes, Solvency II stresses the importance of an "ability to classify risks"[30]. If the

[29] This dispersive factor would be greater if we took into account the calibrations linked to non-proportional reinsurance, but reinsurers tend to systematically develop an internal model. The factor indicated corresponds only to the 9 classes of direct insurance and proportional reinsurance on which this article concentrates.

[30] "Regardless of the calculation method chosen, the ability of the internal model [this article concerns internal models] to rank risk shall be sufficient to ensure that it is widely used in and plays an important role in the system of governance of insurance and reinsurance undertakings, in particular their risk-management system and decision-making processes, and capital allocation in accordance with Article 120." (EPC, 2009), Article 121.4

hierarchy of risks is not captured by the measure being used, then quantification will not improve arbitrations, and possibly degrade them by forcing the substitution of erroneous quantifications for potentially existing alternative analyses.

The objective is to determine to what extent a change in approach modifies the perception of one risk as being more (or less) risky than another. Thus, we compare, two by two, a set of risk pairs in order to determine their relative position in a given approach; then, we observe for a set of approaches whether this comparison is stable by measuring the proportion of pairs whose relative position is independent from the approach used.

The stability measure of the relative position of risks can be formally represented as:

Let $R_{i,j}$ be the risk measure *i* in the approach *j*,

Let $C_{i,i',j} = \begin{cases} 1 \ if \ R_{i,j} > R_{i'j} \\ 0 \ if \ R_{i,j} = R_{i'j} \\ -1 \ if \ R_{i,j} < R_{i'j} \end{cases}$ expressing the relative position of risks *i* and *i'* seen from approach *j*,

Let $T_{i,i'} = \begin{cases} 1 \ if \ sign\ (C_{i,i',j}) = constant \ \forall \ j \\ 0 \ otherwise \end{cases}$

$$Stability = \frac{\sum_i \sum_{i'} T_{i,i'}}{n(n-1)/2}$$

This indicator is between 0 and 1, and represents among the set of possible risk pairs the proportion of those whose risk measure hierarchy remains stable throughout the studies.

Appendix 3: Operation and modalities of the signal-noise ratio adjustment

In this appendix, we formalize and discuss the reasons for and consequences of choosing an adjusted signal-noise ratio rather than a raw signal.

Formalization

We can write the calibration of the risk $R_{i,j}$ associated with the LoB *i* in the set of calibrations (e.g., QIS) *j* in the form

$$R_{i,j} = V_i + n_{i,j}$$

Where V_i represents the value of the risk parameter sought for the LoB *i* and $n_{i,j}$ represents the noise associated with the LoB *i* for the set of measures *j*.

This noise $n_{i,j}$ may comprise:

- LoB-specific components. For example, when one LoB is strategic for public authorities looking to develop this market (its calibration will be revised downward), etc.
- calibration set-specific components. For example, the use of a new methodology that takes into account the possibility of fat-tailed laws that would lead to augmenting all calibrations.
- a residue, corresponding, for example, to when, after a particular event, judgment about a given line of business is revised before any new information can refute said judgment.

As such, it can be modeled from three components. If they interact additively, for example, in the form:

$$n_{i,j} = L\varepsilon_i + Q\sigma_j + B\tau_{i,j}$$

Where ε_i, σ_j and $\tau_{i,j}$ are normalized white noises respectively reflecting the random component:

- for ε, related to the LoB,
- for σ, related to the calibration set
- and for τ, non-specific,

and L, Q, and B represent the amplitude of each of these components[31].

Meaning of a non-adjusted ratio

If we take L and B to be zero, which is to say that we consider the sole source of noise to come from the different sets of measures and that it affects all classes homogeneously, then we would have:

$$R_{i,j} = V_i + Q\sigma_j$$

In this case, our noise indicator corresponds to $\bar{Q}$, and the entire dispersion for a column corresponds to information. Thus, our raw signal indicator, which

[31] Or, more precisely, $n_{i,j} = L_j\varepsilon_i + Q_i\sigma_j + B\tau_{i,j}$, with L_j and Q_i respectively representing the amplitude for each calibration *j* of the noise affecting each class and the amplitude for each class *i* of the noise resulting from the change of QIS.

corresponds to $\overline{dispersion_i(R_{i,j})} = \overline{dispersion\ (V_i)}$, effectively captures the amplitude of the signal.

Thus, the *unadjusted* raw signal-noise ratio, which determines the signal-noise ratio in the case where L and B are zero, is an upper bound of the signal-noise ratio.

This raw ratio corresponds to $1 + SNR_a$ and, applied to the data in this article, it indicates that the amplitude of the noise is of a comparable order of magnitude to that of the signal:

	st. dev	max/min
Standard formula - premium	1.3	2.0
Standard formula - reserving	1.1	1.5
Internal model - reserving	1.7	2.1

Figure 14: raw signal-noise ratios (unadjusted)

Meaning of the adjusted ratio

If, being impossible to estimate L nor to posit a hierarchy between L and Q, we assume that $L = Q$, then with, any B, we can rewrite $n_{i,j}$ in the form $n_{i,j} = N\nu_{i,j}$, where $\nu_{i,j}$ is a normalized white noise. We should then correct the dispersion corresponding to the raw signal indicator with the dispersion induced by the noise of characteristic amplitude N, as measured by the noise indicator. This corresponds to the adjusted signal-noise ratio as presented in the body of the paper.

This result holds, more generally, with any L and $Q \ll B$, where we can thus write $n_{i,j}$ in the form $n_{i,j} \cong N\nu_{i,j}$.

Finally, in the case where $Q \ll L$, our adjustment is insufficient, since the noise measured between different QIS is low in comparison to the noise between LoB mistakenly incorporated into the information in the raw signal measurement. The proposed adjusted signal-noise ratio thus overestimates the information provided relative to the noise.

Appendix 4: Successive risk calibrations proposed by the CEIOPS

	2006 QIS2	2007 QIS3	2008 QIS4	2010 QIS5	2014 Preparatory phase	noise indicators stdev	noise indicators max-min/min
1 Motor, third party liability	5.0%	7.5%	9.0%	10.0%	10.0%	2.1%	1.0
2 Motor, other classes	12.5%	3.0%	9.0%	7.0%	8.0%	3.4%	3.2
3 Marine, aviation and transport	7.5%	5.0%	12.5%	17.0%	15.0%	5.0%	2.4
4 Fire and other damage to property	15.0%	10.0%	10.0%	10.0%	8.0%	2.6%	0.9
5 Third-party liability	10.0%	10.0%	12.5%	15.0%	14.0%	2.3%	0.5
6 Credit and suretyship	25.0%	12.5%	15.0%	21.5%	12.0%	5.8%	1.1
7 Legal expenses	10.0%	10.0%	5.0%	6.5%	7.0%	2.2%	1.0
8 Assistance	15.0%	10.0%	7.5%	5.0%	9.0%	3.7%	2.0
9 Miscellaneous non-life insurance	10.0%	12.5%	11.0%	13.0%	13.0%	1.3%	0.3
raw signal indicators Stdev	5.8%	3.2%	3.0%	5.4%	2.9%		
raw signal indicators max-min/min	4.0	3.2	2,0	3.3	1.1		

Figure 15: calibrations associated with premium risk and associated raw signal and noise indicators

	2006 QIS2	2007 QIS3	2008 QIS4	2010 QIS5	2014 Preparatory phase	noise indicators stdev	max-min/min
1 Motor, third party liability	15.0%	15.0%	12.0%	9.5%	9.0%	2.9%	0.7
2 Motor, other classes	15.0%	7.5%	7.0%	10.0%	8.0%	3.3%	1.1
3 Marine, aviation and transport	7.5%	15.0%	10.0%	14.0%	11.0%	3.0%	1.0
4 Fire and other damage to property	15.0%	12.5%	10.0%	11.0%	10.0%	2.1%	0.5
5 Third-party liability	10.0%	7.5%	15.0%	11.0%	11.0%	2.7%	1.0
6 Credit and suretyship	20.0%	15.0%	15.0%	19.0%	19.0%	2.4%	0.3
7 Legal expenses	20.0%	10.0%	10.0%	9.0%	12.0%	4.5%	1.2
8 Assistance	10.0%	15.0%	10.0%	11.0%	20.0%	4.3%	1.0
9 Miscellaneous non-life insurance	20.0%	10.0%	10.0%	15.0%	20.0%	5.0%	1.0
raw signal indicators Stdev	4.8%	3.3%	2.6%	3.2%	4.9%		
raw signal indicators max-min/min	1.7	1.0	1.1	1.1	1.5		

Figure 16: calibrations associated with reserving risk and associated raw signal and noise indicators

PARI

PROGRAMME DE RECHERCHE
SUR L'APPRÉHENSION DES RISQUES
ET DES INCERTITUDES

PARI, placé sous l'égide de la Fondation Institut Europlace de Finance en partenariat avec l'ENSAE/Excess et Sciences Po, a une double mission de recherche et de diffusion de connaissances.

Elle s'intéresse aux évolutions du secteur de l'assurance qui fait face à une série de ruptures : financière, réglementaire, technologique. Dans ce nouvel environnement, nos anciens outils d'appréhension des risques seront bientôt obsolètes. PARI a ainsi pour objectifs d'identifier leur champ de pertinence et de comprendre leur émergence et leur utilisation.

L'impact de ses travaux se concentre sur trois champs :

- les politiques de régulation prudentielle, l'optimisation de leur design technique et leur appropriation pour le pilotage, dans un contexte où Solvabilité 2 bouleverse les mesures de solvabilité et de rentabilité ;
- les outils d'allocation stratégique d'actifs des investisseurs institutionnels, dans un environnement combinant taux bas et forte volatilité ;
- les solutions d'assurance, à l'heure où le big data déplace l'assureur vers un rôle préventif, créant des attentes de personnalisation des tarifs et de conseil individualisé.

Dans ce cadre, la chaire PARI bénéficie de ressources apportées par Actuaris, la Financière de la Cité, Generali et le Groupe Monceau.

Elle est co-portée par Pierre François, directeur du département de sociologie de Sciences Po et Sylvestre Frezal, directeur à Datastorm, la filiale de valorisation de la recherche de l'ENSAE.

PARTENAIRES